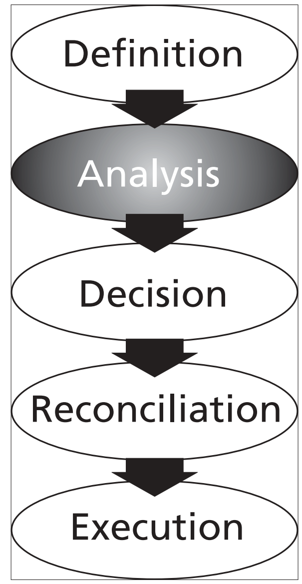
CHAPTER 8

FORCE-ON-FORCE ANALYSIS



With many calculations, one can win; with few one cannot. How much less chance of victory has one who makes none at all!

—Sun Tzu, 400-381 BC, *The Art of War*

FORCE-ON-FORCE ANALYSIS INVOLVES determining how effective a military force is likely to be in combat situations, what factors are most important in determining that effectiveness, and how changes in the force, the adversary, or the combat situation change the likely outcome of combat. Force-on-force analysis is the heart of force structure planning, one objective of which is to develop forces that will prevail in combat. Thus combat effectiveness, expressed in many different ways, is the fundamental criterion we use to compare competing weapons systems, doctrines and operational concepts, force structures, theater Operational Plans (OPLANs), and military strategies.

Force-on-force analysis is the method the U.S. defense community uses to measure combat effectiveness short of committing forces to actual combat. Note that force-on-force analysis is not restricted to questions involving combatant forces only. All military functions, activities, capabilities, and organizations have the ultimate purpose of increasing the effectiveness of U.S. strategic and general purpose combat forces; therefore, we assess them also during force structure analysis in exactly the same way as combat forces themselves.

Formal and relatively abstract force-on-force analysis is a relatively recent invention. Militaries have used exercises in the field for analytical purposes only since the 19th century. The development of indoor force-on-force methods began in the late 19th century in Germany in the form of a board game called Kriegspiel (literally, “wargame”). Dr. Frederick W. Lanchester in Great Britain developed the mathematical roots of force-on-force analysis during and after World War I. In the U.S., formal, institutionalized force-on-force analysis began during the period between the World Wars. Here at the U.S. Naval War College, it took the form of elaborate wargames played at Sims Hall where naval officers developed and tested the amphibious doctrine and aircraft carrier tactics used to win the Pacific War. In the 1950s, mathematical force-on-force analysis using military operations research methods developed during World War II and the Lanchester equations became a basic tool for force planning. Nuclear weapons also lent themselves well to mathematical analysis. The computer has permitted enormously greater elaboration, detail, and speed in the mathematical models available for force-on-force analysis, but as we shall see, computers have not necessarily improved validity.

LANCHESTER'S EQUATIONS

One of the most famous attempts to predict performance of military forces with a dynamic mathematical model was attempted by Dr. Frederick W. Lanchester in 1916. Dr. Lanchester believed that the quantity and quality of military forces determine the outcome of battles and, therefore, that both must be included in any mathematical representation of combat. Briefly, Lanchester asks that we let

A = the numerical (quantitative) factor for force "A"

B = the numerical (quantitative) factor for force "B"

a = the qualitative factor for force "A"

b = the qualitative factor for force "B"

Lanchester postulated that two infantry forces of equal quality, but unequal in size, would inflict casualties on each other based on how many bullets they fired. The larger force would inflict more casualties on the smaller force with each volley; its strength and firepower advantage *relative* to its opponent's would grow after each exchange of fire. This strength advantage would grow as a function of the square of the quantities of soldiers on the two sides, e.g., a force of 2,000 soldiers is *four* times as powerful as a force of 1,000 soldiers. Eventually, the smaller force will be annihilated as it suffered more casualties than the larger force on each exchange while inflicting progressively fewer casualties on its enemy; in this case the force with 2,000 soldiers would take 268 casualties to eliminate the force of 1,000 soldiers. Lanchester named his equation the N-square law because of this attrition phenomenon.

Lanchester also considered how the quality of troops and other environmental factors affected attrition. He ultimately decided that attrition is proportional to the product of the square of the numerical factor multiplied by the qualitative factor of the other force, times a constant "K," which represents battlefield conditions such as weather, terrain, and the like for both "A" and "B." In other words, no single attribute included in K is more important than quantity with its exponential influence. Therefore,

$\mathbf{dA/dt = KbB^2}$ the change (decreased size) of A with respect to time describes the loss rate of force A, and

$\mathbf{dB/dt = KaA^2}$ describes the loss rate of force B

Note that the inclusion of "t" for time is what makes this equation a dynamic representation of combat. Much research has been devoted to testing whether Lanchester's equations and their implications have proven true in combat. In their simplest form, the answer is "No." This is not surprising since Lanchester developed his equation to represent combat as it existed immediately before and during World War I.

Lanchester's equation does a better job of predicting combat results when it is made more complex by adding additional terms representing morale, training, command and control, intelligence, and the like. Through a long process of adding and modifying terms, descendants of the Lanchester equations drive our current generation of force-on-force, campaign-level computer models.

Force-on-force analysis ranges from highly detailed engineering evaluations of individual weapons and their components to much more general assessments of global warfare. Between these poles lie such efforts as evaluations of aircraft, ships, and vehicles; analyses of the organization and effectiveness of tactical units and operational concepts; and assessments of the adequacy of joint theater forces. Force-on-force analysis is performed at every echelon of military decision making. A small unit commander planning an operation uses force-on-force rules of thumb to develop his plan. Similarly, a unified commander preparing theater OPLANs uses field and map exercises, wargames, and mathematical models to test alternatives involving different friendly and adversary forces, courses of action, adversary strategies, and other variables such as weather and terrain.

A military service staff in Washington preparing its annual Program Objective Memorandum analyzes its service's force structure to see if the existing and near-term programmed forces meet projected threats and the nation's obligations for forward deployments. Components of each military service use force structure analysis to develop and test alternative systems for acquisition to equip their operating forces. The Joint Staff performs continuous force-on-force analyses in the Joint Strategic Planning System that creates the nation's overall military direction. The multi-service, multi-CINC, and Joint Staff teams that together make Joint Warfighting Capabilities Assessments use force structure analysis to monitor the ability of U.S. forces and operational concepts to secure U.S. objectives and to warn when those capabilities are falling short. Additionally, the research arms of DoD, such as the service and national laboratories and the Defense Advanced Research Projects Agency, use highly technical force structure analyses focused on proposed weapon technology to determine how much leverage each technology might have on battle situations.

Theory of Combat

Fundamentally, all of our approaches to force-on-force analysis are underpinned by theories of combat that include both how combat works and what matters most in determining the outcomes of engagements, battles, campaigns, and wars. The various analytical methods we use can shed light on the performance of the force alternatives *only* to the extent our theories of combat are valid. If our theories are flawed, our analytical results are likely to be equally wrong. This is why some critics are skeptical of dire global warming predictions. The predictions are based, they say, on climatological models built on a very imperfect theory of how the Earth's climate works. For this reason, it is important to consider how our theories of combat are developed and where they come from.

Combat is an exceedingly complex simultaneous interaction of many factors. Large institutions, here and abroad, have been engaged for many years trying to develop theories of combat powerful and reliable enough to permit accurate predictions of combat outcomes. The most successful theories have been those for predicting the effectiveness of individual systems in combat in which most of the data comes from the physical realm (velocity, penetration, rate of fire, mean time between failures, etc.). We are reasonably confident that we can predict how a particular missile or radar will behave under different operating conditions. Our confidence falls rapidly as we try to forecast the results of more complicated combat situations that depend on the interactions of many weapons and units over time and that are critically shaped by human behavior and decision making. This does not mean that force-on-force analysis is useless for problems other than hardware selection. It does mean that we use the results fully aware of their limits.

CASE STUDY: MAJOR LEAGUE BASEBALL

Suppose you were the manager of a major league baseball team. You would execute a form of force-on-force analysis nearly every day. As you prepare for next season, you analyze the strengths and weaknesses of your team in terms of offense and defense, the pitching staff, fielding, hitting, and base-running ability. You also take into account similar information about your opponents, weighting most heavily the strengths and weaknesses of the other teams within your division. Data abounds, and you may have a considerable chore deciding which statistics to use. But, in general, the data would be reliable and easily available. From your analysis, you develop your trading strategy, shape your minor league teams, and modify your coaching staff.

During the season, play-offs, and throughout the championship series, you assess the opposition in detail and make near-term decisions such as which pitchers start which games and how you will rotate them. During each game, you decide tactics: whether to relieve a tired pitcher, walk a batter, insert a pinch hitter, order a bunt or a sacrifice fly, etc.

At each level—strategic, operational, and tactical—you grapple with the same issues as an executive decision maker in DoD. It's all force-on-force analysis. Which of the many pieces of data are meaningful (valid) and allow you to forecast the future? At what point does your method of force-on-force analysis become so complex and burdensome that its reliability and practicality suffer? You could develop an analytical approach so elaborate that it required the entire season and the team's payroll to run just once. The result might be quite valid but of no use at all.

Force-on-force comparisons cannot accurately predict who is likely to win or lose an engagement, a battle, a campaign, or a war. But we can use them with more confidence (though far from certainty) to predict whether one system, tactic, force structure, or course of action is likely to perform *roughly* better or worse than another. We can also use force-on-force analysis to assess why one alternative performs better than another and what happens to that performance when we change the forces, how they are used, and the conditions in which the combat occurs. But we must always keep in mind that if a force-on-force analysis embodies a completely erroneous theory of combat, even these more modest predictions are likely to be completely erroneous as well. That is why most of the complaints about the adequacy of force-on-force analytical methods, especially those involving mathematical models, are actually about the weaknesses of our theories of combat, especially future combat. This is the argument of many proponents of the Revolution in Military Affairs—that force-on-force analysis, as done today, improperly represents new weapons, technology, and operational concepts, thereby slowing their introduction into operational units.

Methods of Force-on-Force Analysis

Next, we will survey the most common force structure analysis methodologies used in defense resource allocation. They vary in their complexity from very simple order of battle comparisons to highly interactive dynamic models requiring tremendous computing support. Our previous caveat remains in force: computers have made it easier to model much more complicated theories of combat, but they can do no better than the theories of combat that drive them.

STATIC, SYMMETRIC COMPARISONS

Static methods are so-named because they exclude time. They are snapshots of aspects of the combatants we think are predictive of combat results. Usually, we express static measures as numbers and the difference between them or their ratio is taken to represent the superiority of one side over the other. The most straightforward use of static measures is the symmetric comparison, often referred to as a "bean count." Suppose we are Blue force planners tasked with assessing whether Blue forces are sufficient to defend against attack by Orange forces. This is a classic problem for which force-on-force analysis is used. A static symmetric comparison counts Blue tanks against Orange tanks, Blue aircraft against those of Orange, Blue troops against Orange troops, and so forth. To interpret what these comparisons mean for combat, we convert them to ratios. For example, a military rule-of-thumb since the days of Napoleon says that, to carry out a successful ground attack, an attacker must have a 3 to 1 advantage over a defender. For our question of whether Blue is at risk from Orange attack, as long as Blue prevents Orange from attaining a 3:1 advantage, we can defend ourselves against Orange attack. Note the theory of combat embodied in static, symmetric measures: in combat, like forces fight like forces, and the force ratio predicts the outcome.

What are the strengths and weaknesses of this approach? Its greatest strengths are its ease of use and transparency. Anyone is able to clearly see what is being compared and how. The data behind the static measures are usually readily available, and the mathematics involved is usually simple, so the measures are very reliable. Also, we know there is at least an element of truth in the theory of combat they embody. Numbers do matter. But how much do they matter and are they all that matters?

The weaknesses of these measures flow from the weaknesses in the theory that underlies them: quantity is not the only thing that matters in combat. In fact, there are many situations in which quantity may be the least important factor in determining the result of combat. Quality matters a great deal—great numbers of combat ineffective troops are irrelevant. Since World War II, the U.S. has chosen strategies that emphasize precise firepower (quality) over quantity. Logistics matters. So does command and control. Morale or generalship may dominate everything else. Also, modern combat is a combined-arms activity. Tanks fight infantry and anti-tank weapons in addition to other tanks. Artillery engages tanks, infantry, and other artillery as well as anything else in range. To remedy these weaknesses, we must add complexity.

Qualitative Differences

The first improvement we can make in the static, symmetric bean count is to account for the obvious qualitative differences between symmetrically arrayed forces. For example, we can modify the purely quantitative comparison by a multiplier, which represents the relative quality of the forces being compared. Usually we select one weapon or unit as the base (valued at 1.0) and compare the others to it. For example, comparing U.S. and Russian tanks and fighters, we may decide:

1 **M1A1** tank = 1.4 **T-80** tanks and

1 **F-15** fighter = 3 **MiG-25** fighters

These ratios reflect the qualitative edge that superior weapons give a force in combat; in this case we are saying that 1 M1A1 is worth 1.4 T-80s and 1 F-15 is worth 3 MiG-25s. We base these estimates of the relative quality on field data, professional judgment, and empirical evidence from laboratory comparisons. What began as a static, symmetric comparison of Blue versus Orange forces, based on quantity only, can be both quantitative and qualitative. So, if Orange has 300 M1A1 tanks, Blue can buy 101 M1A1 tanks or 141 T-80s (whichever option is less expensive) to prevent Orange from developing a 3:1 advantage.

What are the strengths of this approach? Obviously it represents a more complete theory of combat. Now that we have included quality, the validity of our static force-on-force analysis has improved; we more accurately model the real world. Unfortunately, qualitative comparisons are often subjective and difficult to make. Experts often disagree over the importance of a particular aspect of a system's performance and its proper weight when they establish a quality rating. For example, how important is the top speed of a fighter aircraft? Experts disagree and much depends on how you envision the aircraft will be used, itself an uncertain judgment.

This means that the numerical multipliers representing quality are difficult to agree upon, raising issues of reliability (are we measuring accurately?). How certain can we be that an M1A1 tank is actually 1.4 times better than a T-80 vice 1.3 times better, or twice as good? If F-15s killed MiG-25s at a three to one ratio during Red Flag exercises in Nevada does that mean they will achieve the same results in real combat… or was this kill ratio due in large part to some other factor such as crew proficiency? Our aggressor squadrons are generally far more proficient than our likely adversaries are. We know combat conditions also affect the importance of quality. There is some terrain where an M1A1 is worth at least ten T-80s, e.g., on the defensive with prepared firing positions, at night, and with long, open fields of fire. As we add more quality factors to better reflect the complexity of modern combat and improve validity, reliability declines as we introduce more measurement error.

Intangible Factors

Even if we have properly evaluated the quality and quantity of the weapons on each side, we still have not included some of the major factors that determine combat results. Many military commanders believe training, morale, unit cohesion, leadership, and generalship do more than anything else does to determine the combat effectiveness of a force. How can we incorporate these into static measures? Usually, we can use the same process we use for quality—we can apply a multiplier. The multipliers we use for morale, command and control, logistics, impact of casualties, etc., depend heavily on our theory of combat. An expert panel using whatever data is available chooses a number to capture the intangible capabilities of the two sides. For example, in the 1991 Gulf War with Iraq, many analysts compared U.S. and Iraqi ground forces. Because of many qualitative and intangible differences, the numerically smaller U.S. and coalition forces developed a far greater than 3:1 effective combat power superiority against their Iraqi foes.

The strength of this approach is that it improves the validity of modeling combat by incorporating more of the factors that we believe determine combat outcomes. The weakness is that numerical estimates of intangibles seem even more difficult, dubious, and unreliable than those for quality.

STATIC, ASYMMETRIC COMPARISONS

As we noted earlier, we need to ask whether it is valid to compare tanks against tanks, aircraft against aircraft, and so on, only symmetrically. Or, should we compare them asymmetrically by counting tanks against anti-tank systems and aircraft against air defense systems? Undoubtedly, introducing some sort of asymmetric comparison is appropriate because actual forces do not usually fight in a symmetric manner. In fact, one of our principal tactical objectives is to mass firepower and create a situation where we fight asymmetrically, strength against weakness, by bringing overwhelming combat power to bear at the point of attack, e.g., pitting all of our anti-armor systems— armored fighting vehicles, attack helicopters, close air support and supporting fires—against the enemy armor force.

Figure 8-1 displays an asymmetric theory of combat for World War II weapons systems. The direction of the arrows indicates success, thus armor can attack and defeat infantry, artillery, and air defenses but can be defeated by anti-tank systems and aircraft. Introducing asymmetry surely adds validity to our force-on-force analysis. Unfortunately, just as before, the addition of another factor also introduces greater unreliability. One reason reliability declines is because we have to decide what to compare with what, and how. For example, tanks are simultaneously armor, anti-armor, antipersonnel, fire support, and even anti-aircraft weapons. If we seek to compare Blue armor versus Orange anti-armor, we would surely include the tanks of both sides. But when we compare Blue infantry versus Orange antipersonnel weapons, should we count Orange tanks again? And when we compare Blue helicopters with Orange anti-aircraft weapons, should we count Orange tanks (with air defense machine guns) yet again?

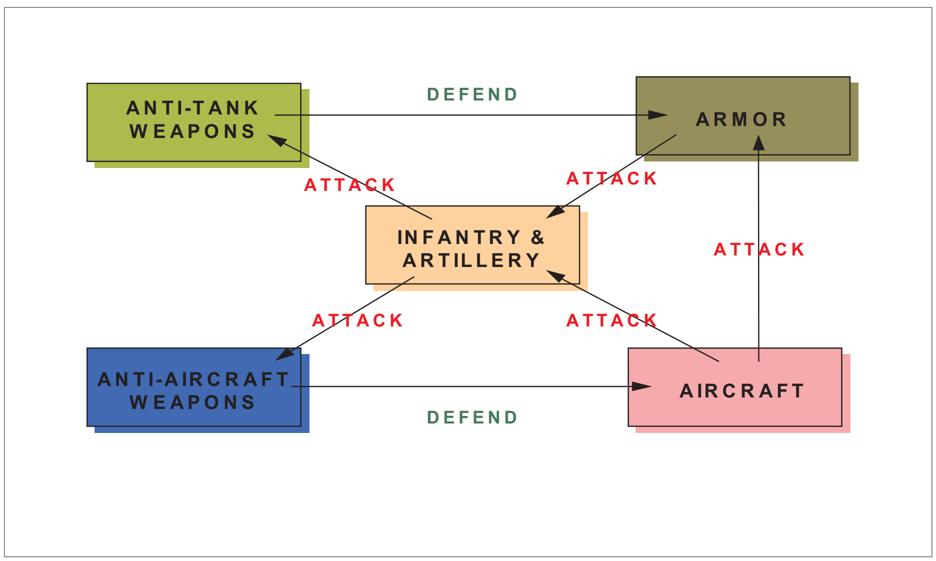


Figure 8-1. World War II Asymmetric Theory of Combat Model[1]

All multi-purpose weapons pose this problem of potential multiple counting. We must decide on the basis of the type of combat we anticipate whether a tank is best in one of its roles or another, or how to apportion it among them. A similar problem arises because only part of each side's force structure engages part of the other's at any given time. We distort this reality when, for example, we compare the total number of Blue tanks to the total number of Orange anti-armor systems. Different people would reach different decisions on these issues; hence this greater complexity introduces more reliability problems. It is difficult to reach consensus among experts about identical asymmetric force-on-force comparisons; as we cannot measure consistently.

SUMMARY OF STATIC COMPARISONS

We have seen how static measures can provide simple, clear snapshots of military capability at the price of limited validity: too much of what matters in combat is excluded. The remedy is to

1. Adapted from Archer Jones, *The Art of War in the Western World,* (New York: Oxford University Press, 1987), p. 611.

CASE STUDY: THE U.S. ARMY AND WEI/WUV'S

In the mid 1970's, the U.S. Army began a project to measure the difference in combat power among different types of Army divisions. Its earlier attempts focused on counting manpower in terms of Manpower Division Equivalents or MDEs. One MDE was a collection of 18,600 personnel in uniform. This calculation, easy to perform and hence reliable, was low in validity because it did not distinguish among types of equipment or tactical mobility, which are obvious determinants of combat power. Many leaders and analysts saw little utility in using MDEs as the basis for static comparisons. Next, the Army created a new unit of comparison called the Armored Division Equivalent, incorporating both manpower and quantity of weapon systems, although quality was still excluded. For this reason, the Army next developed a method for incorporating quantity and quality into static comparisons of ground forces.

The first step was to calculate, through laboratory and field testing and by convening panels of military professionals, a firepower score for each weapon in the Army inventory. These scores represented the accuracy and killing potential for every weapon against a standard target at a fixed range. Similar scores were developed to represent every weapon's mobility and vulnerability. By combining these scores, the analysts extrapolated a general capability index for each Army weapon. The result was a set of Weighted Equipment Indices, or WEIs, for each rifle, tank, cannon, etc.

Given WEIs for each weapon in a division, and the table of organization and equipment for each division, the analysts summed the WEIs to derive a single numerical score, called a Weighted Unit Value (WUV), for each U.S. division, our NATO allies, and our Warsaw Pact adversaries. For convenience, the score of the 2nd Armored Division, the heaviest and most powerful armored division in the world at the time, was given a score of 1.0 and other U.S. and foreign unit scores were set relative to it, e.g., the most capable Soviet Armored Division was given a score of 0.8.

The WEI/WUV methodology was very useful for comparing the combat effectiveness of units. While it was still a static measure, it was created using estimates of how units would perform in combat in terms of firepower, mobility, and vulnerability. Thus, it addressed some of the most important validity problems that afflict most static measures. The WEI/WUV method became a standard in DoD analytical communities and analysts used it as a starting point for wargames and umpiring battle results.

Unfortunately, the WEI/WUV approach had shortcomings. For instance, there was no WEI/WUV equivalent for close air support, yet air forces have a pronounced effect on the ground battle. Also, simply summing the capabilities of a unit's weapons did not capture the synergistic effects of the weapons operating together. That synergy, called combined arms by the Army, is central to operational concepts for conventional forces. Finally, intangibles such as training, morale, and leadership were not included in the WEI/WUVs. As a result, the Army has abandoned the WEI/WUV method in recent years as a force planning tool, although similar indices are still used as components in larger analyses.

include more of those missing factors at the price of decreasing reliability and practicality. As we add more complexity to static comparisons, we have seen how it becomes increasingly unclear exactly what the numbers mean that we seek to compare. When we began by comparing only quantity, the meaning of the numbers was completely clear. They represented the size of two forces. But each time we modify our numbers by including quality, intangible factors like morale and training, and then the asymmetry of combined arms, the harder it is to explain to ourselves and others what those once simple numerical comparisons actually represent and why our models work the way they do.

As we gain in validity, we lose the reliability that is the principal strength of static measures; we also lose transparency and reduce the ease of understanding the model, particularly when we are dealing with the uninitiated. The consequences of these weaknesses are severe. Either we give up trying to represent much of what we think matters in warfare (loss of validity), or we have to make the static measures increasingly complex, opaque, and arbitrary (loss of reliability and often practicality). Note that many of the factors that are hardest to incorporate in static comparisons are the very ones most important to us as we assess issues surrounding the Revolution in Military Affairs; e.g., the value and effects of very fast, fully integrated command, control, communications, computers, intelligence, surveillance and reconnaissance; increasingly speedy tactical decision making; and network-centered warfare.

Despite their weaknesses, the relative simplicity of static measures makes them attractive, and we still use them in DoD in many different forms. As you develop, use, and evaluate them, keep in mind that all the various methods of force structure analysis have strengths and weaknesses. None of them produce answers that are completely correct. Sometimes, maybe often, the validity problems that simple static measures suffer are tolerable. It depends entirely on the problem we are solving or the decision we are making. The speed and clarity of well-done static models may more than compensate for their limitations, e.g., they are still very useful for comparing nuclear arsenals.

DYNAMIC FORCE-ON-FORCE MODELS

The most fundamental limitation of static measures is that to remain simple they must exclude time. They freeze the capabilities of a military force at a particular moment. Yet we know that time and tempo are central to military operations, especially as we move closer to *Joint Vision 2020* operational concepts such as precision engagement and dominant maneuver. All the factors that static measures freeze change continually as military operations proceed. As forces maneuver, their new locations often change their capabilities. Similarly, their capabilities also change as they are reinforced and re-supplied, suffer attrition, expend munitions, and move in and out of contact with higher command. Successful military commanders use time more effectively than their adversaries do. Thus, no matter how sophisticated static measures become, they always (by definition) exclude time, a basic factor that determines the outcome of combat. Dynamic measures attack this problem.

Making a comparison dynamic means we incorporate the dimension of time. By better approximating the actual conditions of war, which are certainly dynamic, we increase the validity of the force-on-force analysis. But, as before, this added validity comes at the expense of reliability and practicality—which the use of computers has not overcome.

There are three general approaches toward making dynamic comparisons. The first type, mathematical models and computer simulations of combat, includes time in its mathematical representations of the forces and the variables at play. The second group, exercises and experiments, incorporates time by using the real forces themselves. The last, wargames, models time by having participants play the game in turns while using maps and demarcations, symbols in place of actual forces, and rules governing their behavior. Remember, just as with static models, whatever form dynamic force-on-force analysis takes and no matter how many factors are included, the results can never be any better than our understanding of what determines the outcomes of battle in the real world, our theory of combat. If it's true that dynamic model results cannot be better than the theory of combat embodied in that model, regardless of its computational power, what can we learn from such models? Sometimes we can learn a lot. To see this point, let's consider the case of computer-driven dynamic models.

First, think about what people do well, and what computers do well. Computers are good at keeping track of large volumes of input data, and at consistently following the rules programmed into them. They will tell you exactly what is implied by input data, given all the rules you have programmed in about how to manipulate that data. And they will do so even when any person would recognize right away the results don't "make sense." In contrast, people are good at seeing the big picture, placing facts in context, and being imaginative. On the other hand, human beings are not always consistent—we might say that we believe X, Y, and Z, but nonetheless shy away from implications that don't "make sense" or are otherwise unacceptable.

Given these relative strengths and weaknesses, we can learn from dynamic computer models even though we are the source of everything that those models "know." This is especially the case if the model produces results that aren't what we expected. In that case, the model is telling us that, if we believe input data values A through Z, and if we believe rules "A" through "Z" for manipulating those values do a good job of stating our theory of combat, then we ought logically to accept the model's conclusions. If we don't, it doesn't mean we are wrong and the model's right. (Far from it.) But it should force us to think. Is there something wrong with the input data or the manipulation rules that embody what we said was our theory of combat? Or, given that data and those rules, ought we to accept the model's results?

Other kinds of dynamic modeling can also force us to think. If exercises, experiments, and wargames are honestly run, with players given free rein to do what they see fit, then these dynamic simulations can also produce unexpected results. In trying to explain those results, we can gain insights that help us question the "conventional wisdom" and refine our theories of war. Below, we discuss each type of dynamic simulation in turn.

Mathematical Models and Computer Simulations

Mathematical force-on-force models have grown increasingly sophisticated as computing power has become more available. Generally, the more abstract arrangements tend to be referred to as mathematical models, while the more detailed and complex assemblages are called simulations because they are supposedly more life-like or closer to experiential reality. These models range from very rigorous engineering representations of individual items, through system (often vehicle) simulators that we often network to one another, to more aggregated but still highly complex models of theater military operations. Theater-level combat models almost always play an important role in major DoD resource allocation decisions because such decisions are ultimately aimed at improving combat effectiveness.

Although mathematical models are used to assess all levels and types of warfare, we will focus on the strengths and weaknesses of the theater-level models used by the Department of Defense. They are, at heart, elaborate pieces of software that contain mathematical representations of the aspects of theater operations their designers deemed important to determining theater campaign outcomes. These include air, ground, and sea forces; logistics; weapons of mass destruction; command, control, communications, and intelligence; morale; and strategic lift. In theater-level combat models, the areas of operation are represented as maps with icons representing forces and units. Operations are conducted in blocks of time, sometimes variable. These models usually can be run fully automatically but they also can be stopped at any point, reversed, modified, and rerun, making them ideal tools for exploring branches and sequels and for sensitivity analysis.

In any theater-level combat model, all the air, ground, and sea units and their individual characteristics have to be loaded into the computer, unit by unit, before we can use the model. These data are inserted into an enormous series of spreadsheets. Ground units are usually represented at the brigade level, air units by squadrons or wings, and sea units by battle group, although most models permit operators to use greater or lesser aggregations. Each side's concepts of operations or decision logic must be loaded into the model so that every unit reacts to each eventuality. Not surprisingly, these instructions require constant adjustments since the range of eventualities is so great. In sum, the preparation of the model for use is labor-intensive and full of opportunities for errors that are discovered only by trial and error—if at all.

The most widely used theater-level combat model in DoD is the Tactical Warfare Model (TACWAR), managed for the Joint Staff by the U.S. Army Training and Doctrine Command, although each of the services also has its own models. The Office of the Secretary of Defense and the Joint Staff use TACWAR extensively to examine force planning options. The unified commanders test their Operational Plans using TACWAR. We will use TACWAR, as DoD used it during the Deep Attack Weapons Mix Study (DAWMS) in 1998, as an example for this section of the text.

AIRCRAFT	MK-82	CBU-87	AGM-65	GBU-27	JDAM	JSOW
AV-8B	6	4	2		4	
B-2	80	24			16	16
F-16C	6	4	4		4	2
F/A-18C USMC	12	8	8		4	2
F-117				2		

Table 8-1. Allied Aircraft Weapon Payloads.

Table 8-1 is a partial representation of some of the combinations of aircraft and payloads analysts inserted into TACWAR preparing for the Deep Attack Weapons Mix Study. Note the assumptions the analysts have to make to create a spreadsheet like this: we know these aircraft are capable of carrying these loads, but realistically would they? Are these representative loads for a typical mission in this theater? We know that if the aircraft is based closer to the target, or does not need to loiter waiting for a call from ground forces to strike, it can carry more bombs and less fuel; what did the analysts assume for these aircraft? Also, they do not allow mixing weapons types, etc. We use the data in these tables to represent all the aircraft of that type in the simulation, so the difference between loading four versus six bombs on an AV-8 has important implications for the "worth" of that aircraft and how it contributes to building combat power.

AIRCRAFT	SURGE SORTIE RATE	SUSTAINED SORTIE RATE
AV-8B	3.5	2.53
B-2	0.8	0.55
F-16C	2.5	1.96
F/A-18C USMC	2.9	2.24
F-117	1	0.57

Table 8-2. Allied Aircraft Sortie Generation Rates.

Combined with table 8-1, table 8-2 gives us a clearer sense of how TACWAR works and its theory of combat. The payload and number of missions each aircraft flies (a sortie is one mission flown by one aircraft) together generate combat power; the most powerful aircraft carry more weapons and fly more often. The surge sortie rate is the maximum possible number of missions in a 12-hour TACWAR cycle the aircraft can fly during an emergency, e.g., during the Halt Phase when a breakthrough or overrun of friendly forces or key terrain seems imminent. The sustained sortie rate can be maintained more or less indefinitely.

AIRCRAFT	CLOSE AIR SUPPORT	STRATEGIC TARGETS	SAM SUPPRESSION	GROUND-CON TROLLED INTERCEPT SITES	AIR BASE ATTACKS
AV-8B	1.00				
B-2	.6	.2	.2		
F-16C	.3	.05	.4	.15	.1
F/A-18C USMC	.45	.2	.25	.1	
F-117		.4	.3	.3	

Table 8-3. Allied Aircraft Target Allocations.

TACWAR needs to know the missions and target sets the planners will allocate to each type of aircraft. Table 8-3 shows the distributions of effort for the aircraft from the earlier tables that were used in DAWMS. For example, 60 percent of B-2 missions will strike enemy ground forces invading the nation we are defending, 20 percent will attack strategic targets like power grids and command centers, and 20 percent will attack enemy Surface-to-Air Missile batteries. To build this table, the military planners must make operational choices about the overall air campaign for the theater and then set the level of effort for each different target set. The planners must also decide whether they will change their apportionment during different phases of the campaign. TACWAR has five Attack Mission categories, some with as many as four sub-categories, and two Defense Mission categories (Battlefield Defense and Area Defense).

Just as we discussed earlier with static combined-arms models, multi-mission aircraft pose a problem for the analysts. Their roles may actually be situational, dependent on enemy actions and levels of activity. Analysts can try to accommodate these actions in the model with a series of "If… then" rules, but they do so by introducing yet more complexity and they require extensive help from operators to ensure they use reasonable rules.

Apportioning aerial effort is where service cultures and doctrines clashed so mightily during the Gulf War: does the CINC or his/her J3 make this apportionment decision or does the Joint Forces Air Component Commander (JFACC)?[2] How do we account for excess sorties the Navy and Marines will provide after they have met their own requirements? How many sorties of

2. Current joint doctrine has the Joint Forces Air Component Commander propose the apportionment to the CINC. This allows the other Component Commanders a built-in opportunity for reclama if they disagree with the proposed apportionment.

what type will halt an aggressor without sacrificing key terrain? What if resources are limited and the CINC needs more sorties than the maritime services volunteer? While many of these answers are rooted in doctrine and service procedures, TACWAR can tell us what answers we ought to be prepared to accept, if we agree with the data values and other assumptions input into the model. For example, in an otherwise fixed scenario, what happens if a Navy air wing reduces its counter-air defense combat air patrols (interceptors) and dedicates them to ground support? Is the enemy's halt line significantly altered? Does the carrier or its escorts become unacceptably vulnerable?

Another important preliminary step the analysts must set into TACWAR is a map of the theater of operations. After the map boundaries are set based on the scenario we are examining and the physical terrain is input and verified, the analysts identify key military terrain and facilities that affect both sides, e.g., aerial and sea ports of debarkation, roads, bridges, urban centers, bases, economic objectives, etc., as shown in figure 8-2.[3]

Theater-level combat models conduct their campaigns by moving forces to their objectives; along the way they may make contact with the enemy. Generally, ground units advance along scripted axes. As shown in figure 8-3, the sectors (or cylinders) vary in shape and size depending upon how they conform to terrain. The analysts often place more numerous, smaller sectors in areas where they anticipate contact between opposing forces.

3. The TACWAR figures presented here are labeled as notional because they represent a hypothetical (illustrative) scenario versus one prepared or confirmed by intelligence officers, logisticians, and other staff planners for actual operational planning.

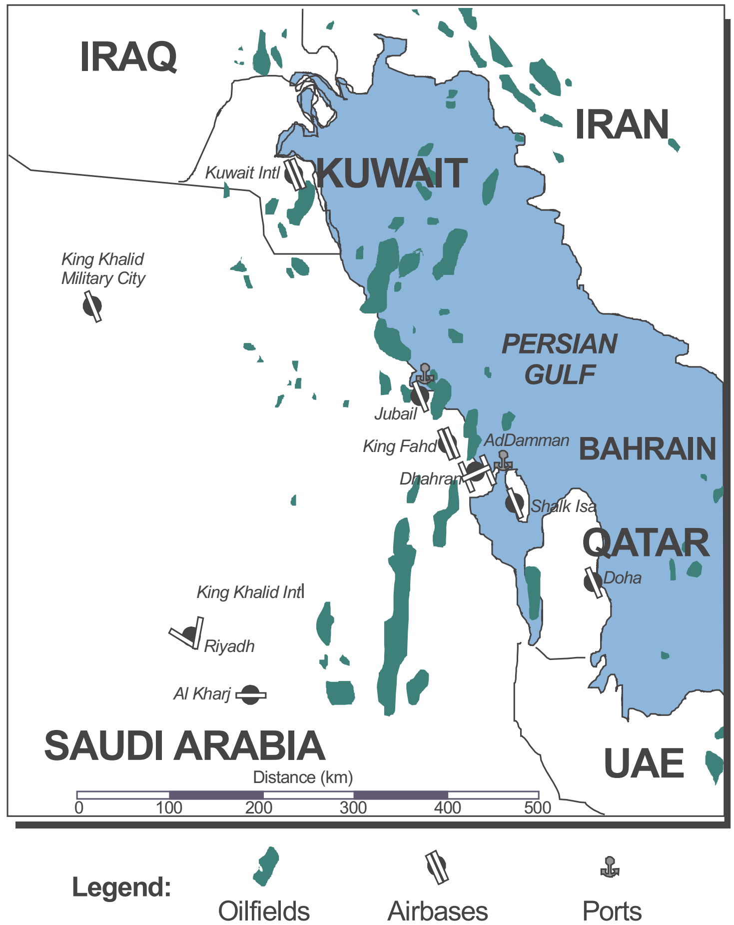


Figure 8-2. Notional Key Terrain in Southwest Asia.

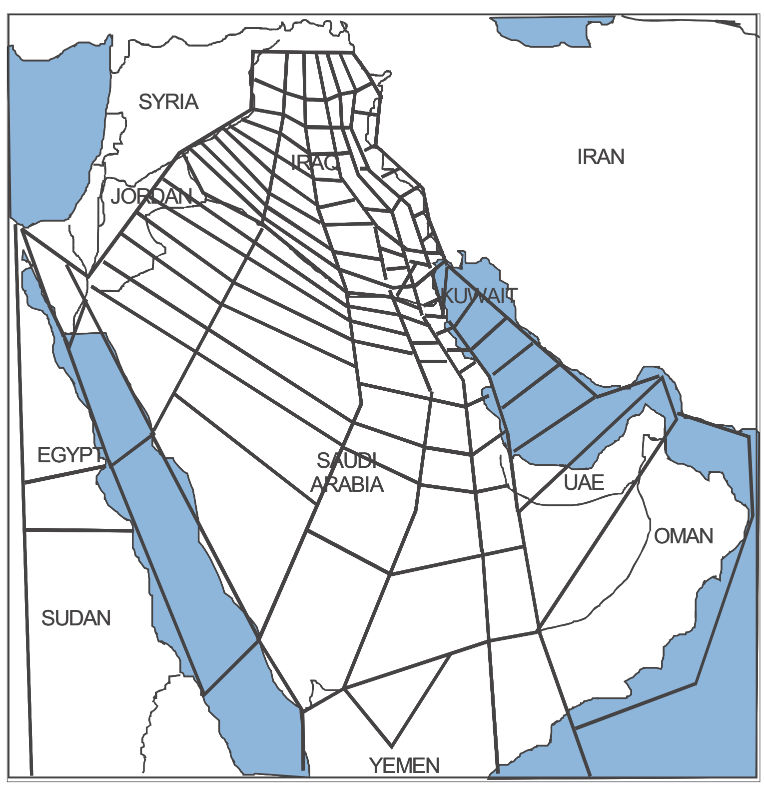


Figure 8-3. Notional Sectors Overlaid on a TACWAR Map.

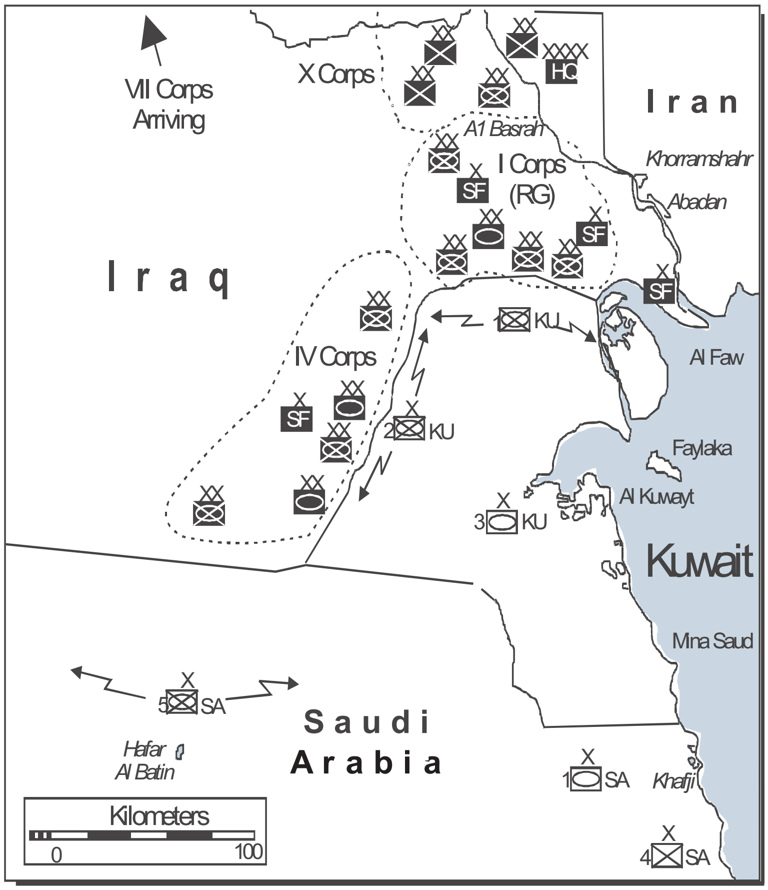


Figure 8-4. Notational Disposition of Allied Ground Forces

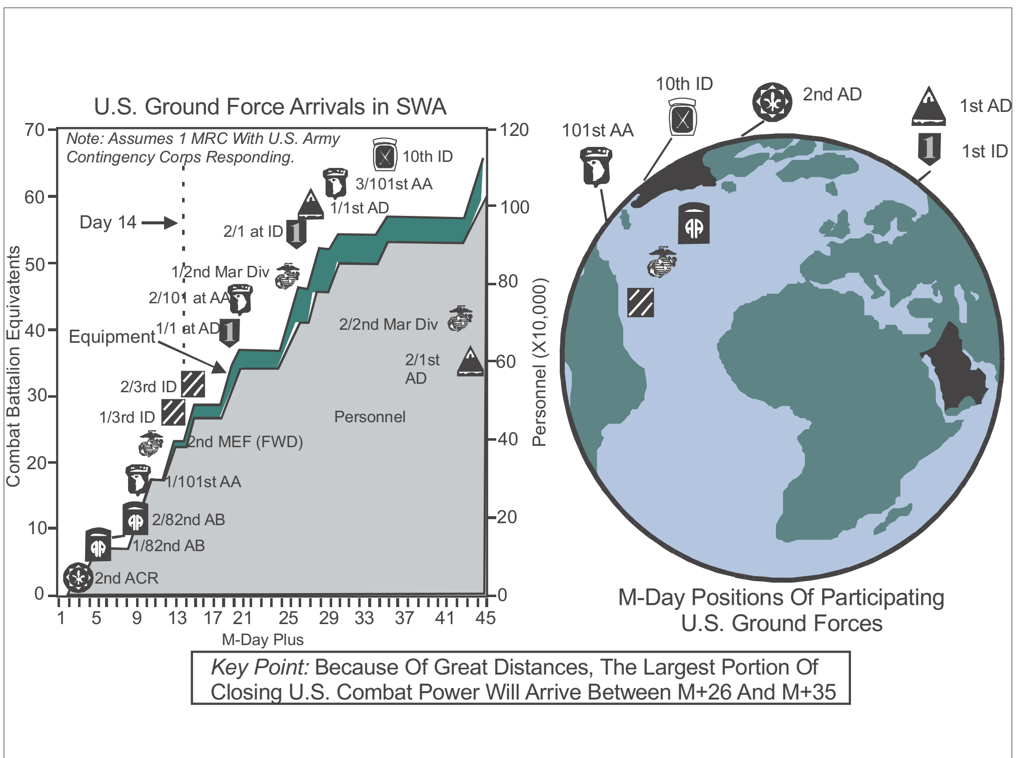


Figure 8-5. Notional Arrival of U.S. Ground Forces.

There are other imaginary lines in TACWAR, too, that the analysts can adjust, e.g., when does close air support of ground troops stop being close air support and become battlefield interdiction? How far behind the battle lines can deep strikes reach?

Now that the map is set and the analysts have specified the capabilities of individual units, they create the initial disposition and flow of forces into the theater. These are variables that may be central aspects of the analysis. For example, how much earlier is the enemy halted (according to the model) if we have three, rather than two, Army brigade sets of equipment pre-positioned in Southwest Asia? The Joint Staff Mobility Requirements Study 2005 used this kind of TACWAR modeling to find the smallest, latest arriving series of forces that halted the invaders short of key terrain with moderate risk. Figure 8-4 shows the initial disposition of Allied forces used in DAWMS for its Southwest Asia scenarios.

Figure 8-5 displays the flow of U.S. ground forces, listed by brigade and measured in battalion equivalents, into Southwest Asia in the Deep Attack Weapons Mix Study. TACWAR allows the strategic lift of ground units to be separated into personnel and equipment because they arrive by different means, airlift and sealift respectively. As used in DAWMS, TACWAR assumes units are ready to move and standing by for strategic lift, i.e., it ignores complicating issues like disengaging from current operations like Bosnia or Kosovo, reconstitution, training, fort-to-port movement, and (for DAWMS) en route attrition.

TACWAR analysts must go through a similar process for aircraft, including arrival times and especially bed down in theater. Figure 8-6 shows the build-up of U.S. airpower from mobilization day forward by aircraft type. Note the rapid availability of long-range bomber aircraft and the steep ramps upward as each aircraft carrier arrives.

All the land-based aircraft that flow into the theater must be bedded down at air bases with sufficient capacity, as we show in figure 8-7. Here is where coalition planning is especially important. While all national planners know the Maximum On Ground capacity of each base, they must coordinate to ensure that collectively they do not exceed it.

Having identified the capabilities of individual weapons and units, and now their quantity as a function of time, i.e., initial dispositions and reinforcements,

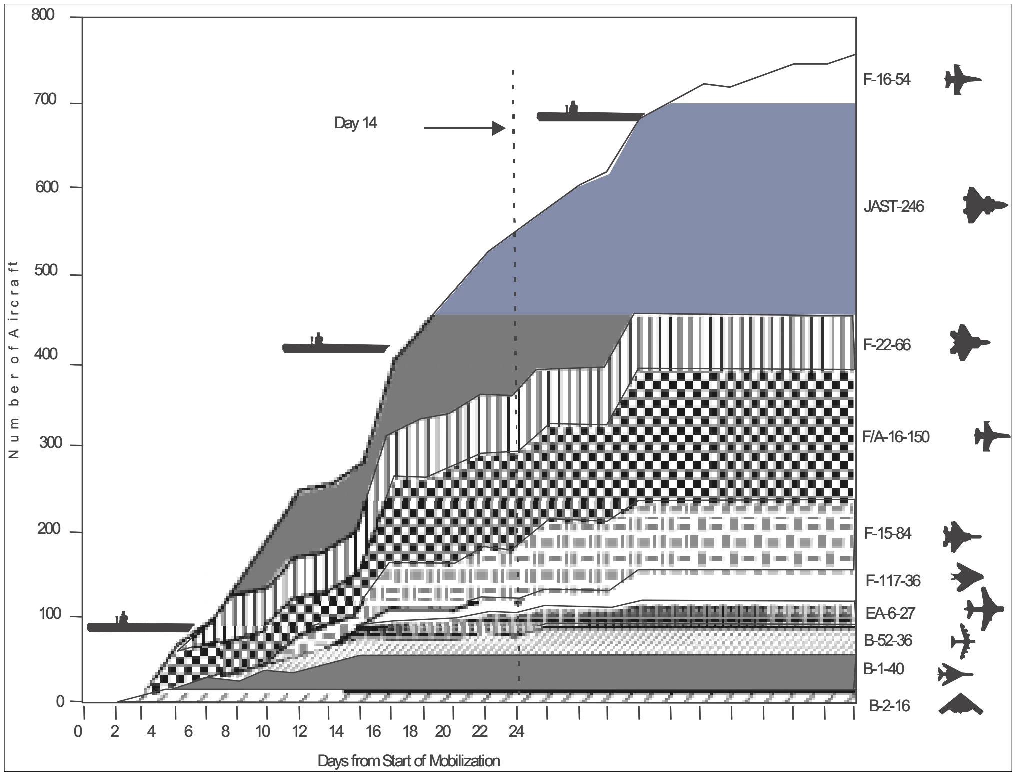


Figure 8-6. Notational Arrival of U.S. Air Forces.

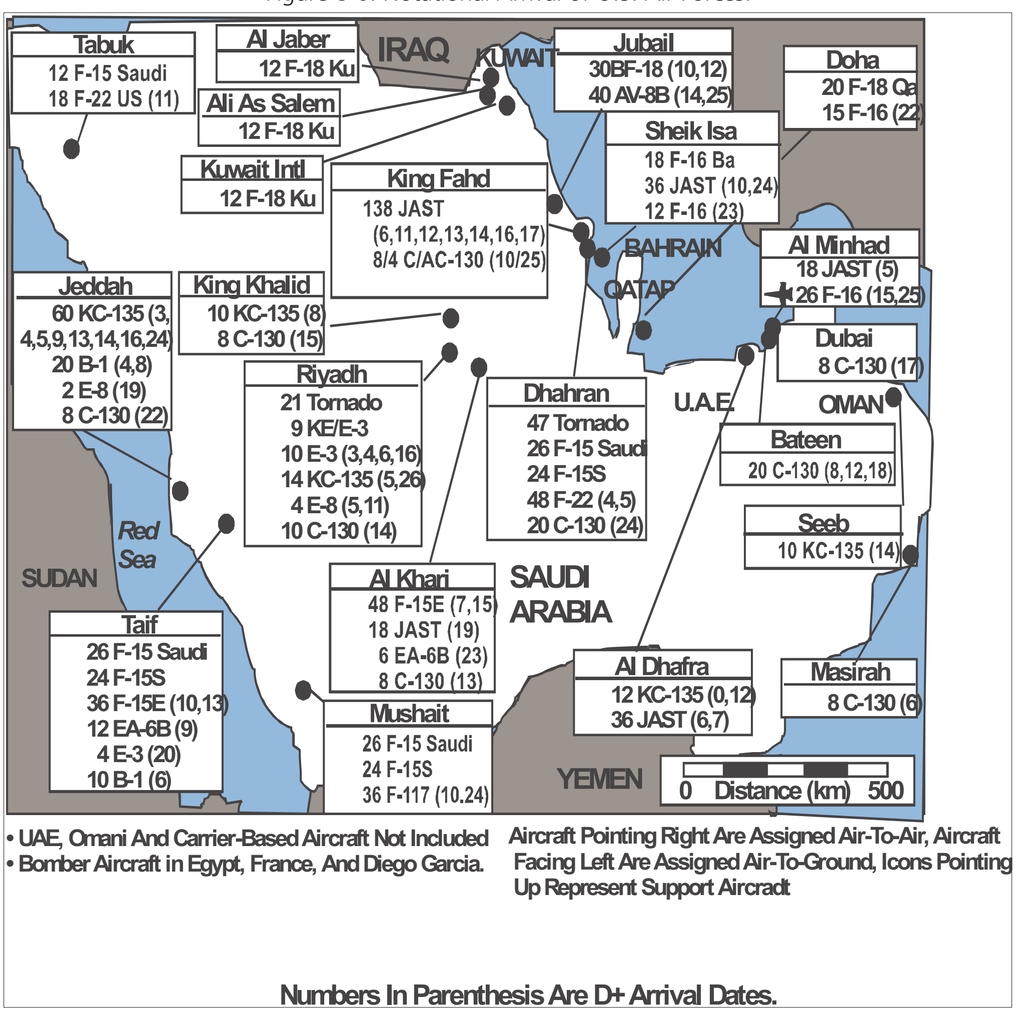


Figure 8-7. Notional Allied Air Basing in Southwest Asia.

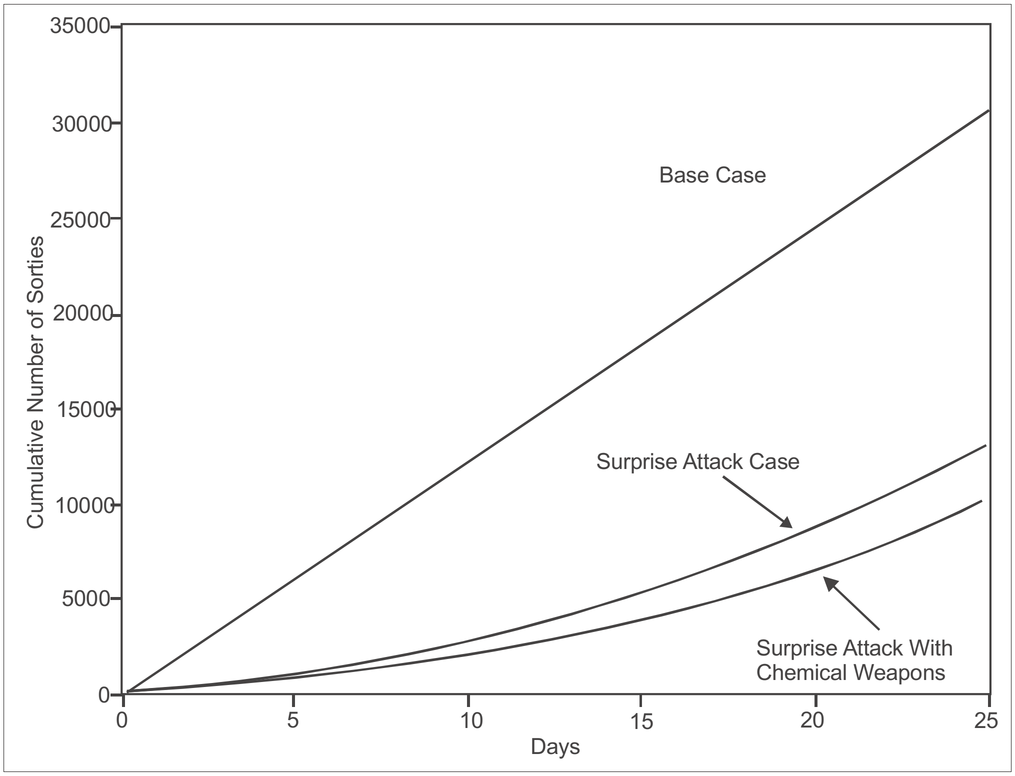


Figure 8-8. Notational Allied Sortie Generation

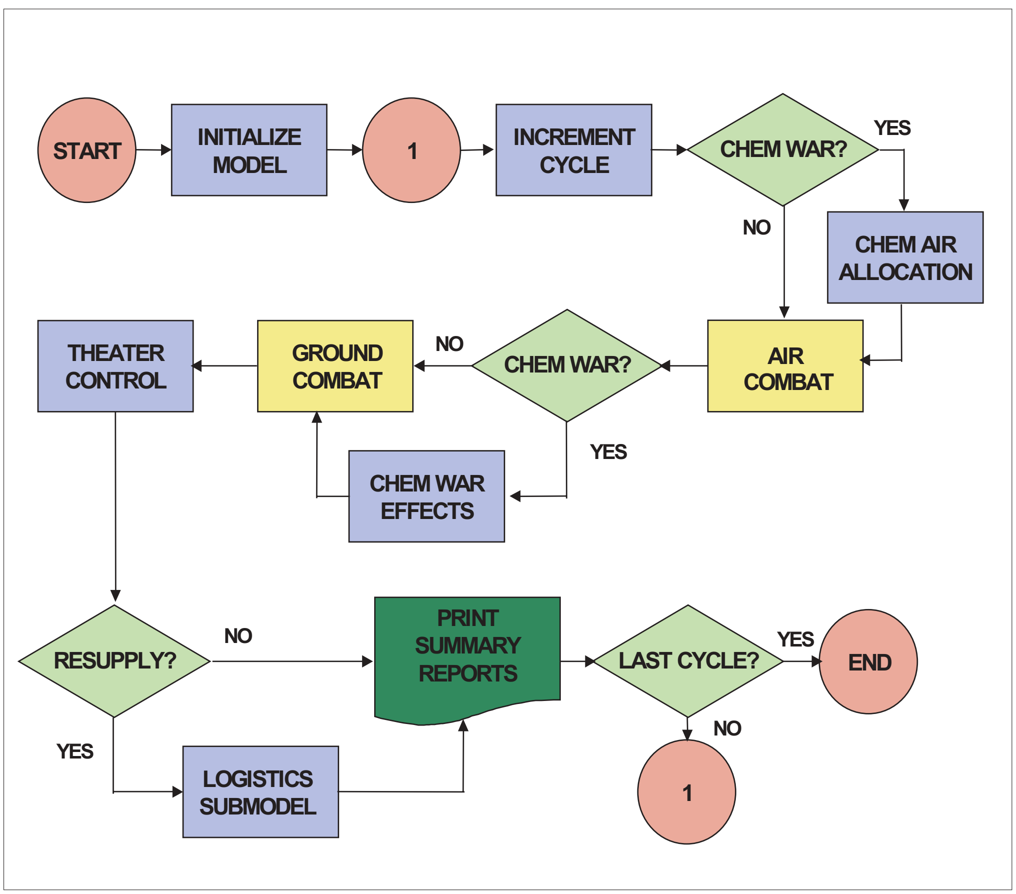


Figure 8-9. TACWAR Model Flowchart[4]

the analysts using TACWAR can calculate how quickly combat power builds up. They can generate charts like figure 8-8 to display aircraft sortie generation and display the sensitivity of this variable to changes in assumptions. The hypothetical base case in DAWMS assumed the Coalition Partners would have advanced warning of an Iraqi attack and that they would refrain from using chemical weapons; this graph shows the impact —according to TACWAR's theory of combat—of a surprise attack and the use of chemical weapons on sortie generation.

The modelers perform a similar analysis of the opposing forces—their initial dispositions, capabilities, likely axes of advance, reinforcement rates, etc., and then they are ready to run the model. TACWAR advances the ground forces along the cylinders until they have moved as far as they can in the cycle or until they make enemy contact.

Figure 8-9 shows the sequencing in TACWAR during each simulated 12-hour cycle. First, the model assesses the effects of the optional Chemical Warfare module and applies modifiers prior to air

4. Based on Figure 0-2 from Steve Kirin's "Executive Summary," *TACWAR Integrated Environment* (U.S. Army Training and Doctrine Command Analysis Center).

combat. Air operations are handled abstractly but within the model and prior to resolving ground combat, which may also be affected by chemical warfare. In TACWAR, players move ships toward and into operating areas or allocate them to patrol routes, but naval movement and combat is handled off line, outside the model. Naval forces inject firepower into sectors much like strategic air power, i.e., without the limitations and vulnerabilities of on-map basing.

TACWAR calculates air superiority within each cylinder. Based on the instructions the analysts provided during the set up, TACWAR assigns counterair units to cylinders in which their strength and capability is compared to the enemy counterair presence. Doctrinally, behind friendly lines, this is Defensive Counterair; in front of them it is Offensive Counterair. TACWAR calculates losses for each side and leaves the residual counterair capability of the superior force in the cylinder for the rest of the cycle. This residual counterair force may have the opportunity to engage strikes and ground support aircraft (and their escorts) in their cylinder based upon their remaining weapons.

Strike and bomber air units "fly" to their targets through the cylinders and, depending on their mission and profile, they may be subjected to attrition from surviving enemy counterair—interceptors and surface-to-air missiles. Support and escorting aircraft such as fighters and electronic jamming aircraft may negate some or all of enemy air defenses. TACWAR then calculates the surviving combat power's effect on their target sets and the users can request TACWAR results as we show in figure 8-10. The graphs indicate the Coalition's reduction of Iraqi Ground-Controlled Intercept and Surface-to-Air Missile battery air defenses over time in the base scenario and two sensitivity variations.

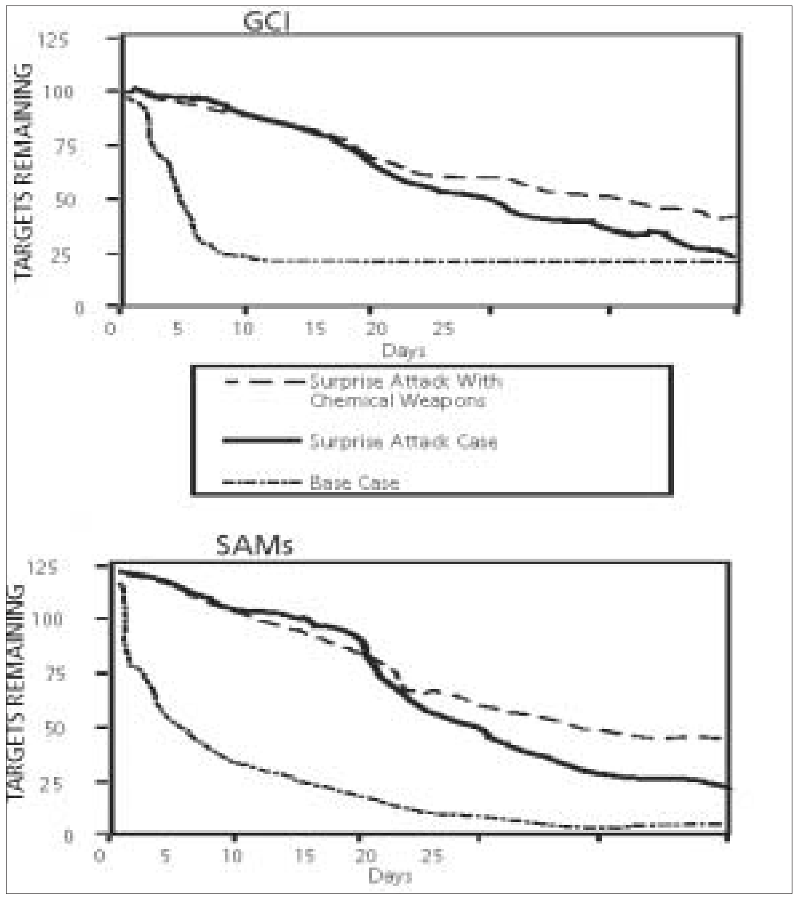


Figure 8-10. TACWAR Target Attrition.

The ground forces and their interactions are the original design focal point of TACWAR; many of its features such as expanded air warfare, logistics, and chemical weapons were added later to improve its validity at modeling modern warfare. TACWAR moves units, has them make various kinds of attacks, or conduct various kinds of defenses based on the instructions of the analyst and the participants. When opposing forces occupy the same cylinder, TACWAR calculates whether either side has enough force superiority to attack (regardless of the overall tactical situation) and, if that side is ordered to assault, it compares the combat power of the forces, including close air support, *in each sector individually*. The threshold required to attack is set by the analyst and may be adjusted between cycles and varied for each side.

The tactical posture of the units affects combat results, e.g., units ordered to delay will fall back to minimize casualties, moving the Forward Edge of the Battle Area (FEBA) with

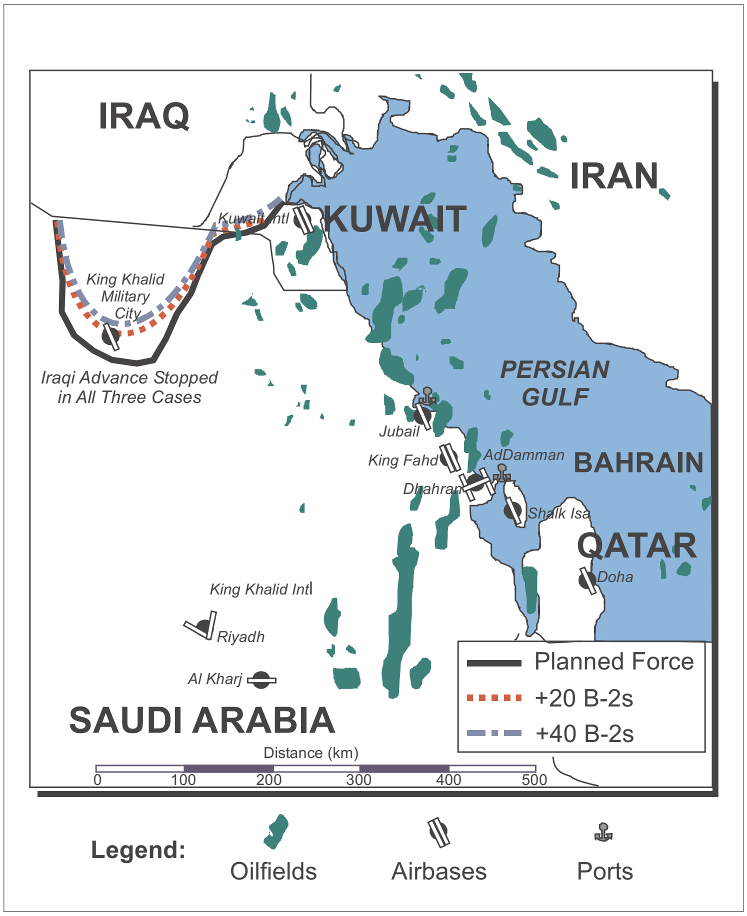


Figure 8-11. Battle Lines Calculated by TACWAR for DAWMS

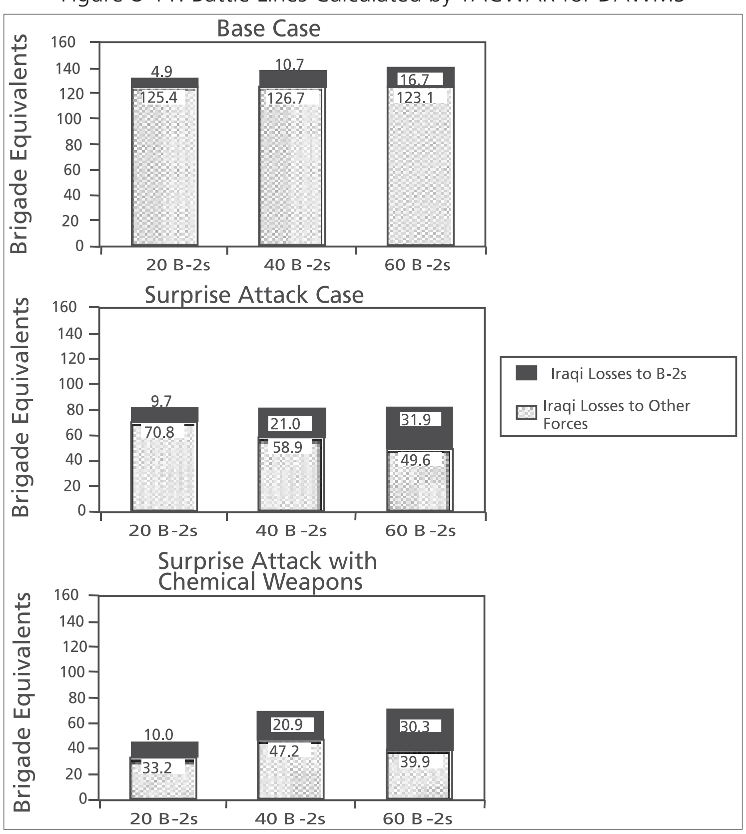


Figure 8-12. Notional Iraqi Ground Force Losses.

them as they fall back, whereas a unit ordered to hold will take greater casualties to prevent the capture of an objective and prevent movement of the FEBA. Depending on the combat results, the battle line between the opposing forces moves back and forth along the cylinder axis like a piston. Hence, models like TACWAR are often referred to as "piston-driven" models. TACWAR then links together and smoothes the piston positions of each cylinder to display the theater battle line after each cycle as shown in figure 8-11.

Note that figure 8-11 contains some sensitivity analysis. The Deep Attack Weapons Mix Study had a two-fold purpose: first, to identify the optimal mix and quantity of deep attack weapons among services and, second, to determine whether the U.S. should purchase additional B-2 bombers so as to place the order for more aircraft before the production line shut down. The different battle lines reflect the contribution of one and two additional increments of 20 B-2 bombers.

In order to establish how the battle line shifts, the combat resolution table determines the results of each enemy contact at the end of each time block in terms of casualties, logistics consumed, ground gained or lost, and targets destroyed. Those figures can be extracted after any cycle in the scenario, or at its conclusion as shown in figure 8-12 that shows Iraqi losses. These graphs are taken from the second part of DAWMS and reflect sensitivity analysis in both the quantity of B-2s and in the nature of the scenario. They demonstrate how tactical surprise and chemical weapons reduce allied effectiveness (fewer Iraqi losses) and how B-2s are relatively unaffected by either.

Figure 8-13 displays the reverse of the coin in figure 8-12, U.S. and Coalition Partner ground losses. They further demonstrate, as B-2 advocates would argue, that B-2s are increasingly valuable as the situation becomes more dire for the allies—nations who are becoming increasingly sensitive to casualties.

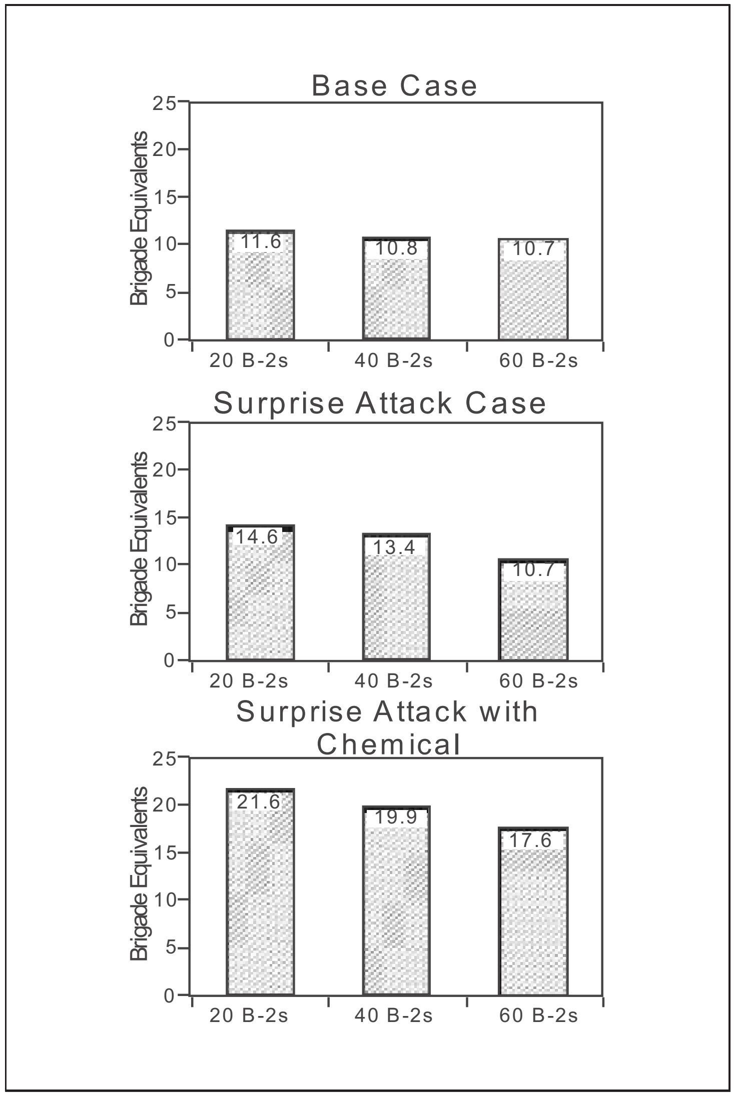


Figure 8-13. Notional Coalition Ground Force Losses.

The combat resolution table is one of the key components of any theater-level combat model. For TACWAR and the other theater-level models we use today, the combat result engines are collections of mathematical representations (equations and matrices), modified forms of Lanchester's equations adjusted with combat data gathered mostly from World War II and the Korean War. TACWAR begins combat resolution by aggregating the strength and quality of the opposing sides' weapons (much like the WEI/WUVs we discussed in our earlier section on static force-on-force methods) to calculate attrition in each sector.[5] It adjusts each unit's weapons' effectiveness based upon adjustments for logistics, training, chemical weapons, tactical posture, etc., and determines how many enemy personnel became casualties, how many enemy weapons were destroyed, how much ammunition and fuel was consumed during the 12-hour exchange of fire, and, as a result, how the FEBA shifted.

TACWAR also has several other modules that can be turned on or off or used to change data during an analysis. The Logistics Submodel overlays a network of supply points and places a hierarchical distribution grid over the sectors. Both are vulnerable to air attack. It monitors consumption and resupplies units in the priority order set by the analyst. The Theater Control routines control the interface with the map and its sectors that affect the level of detail or granularity in the scenario, i.e., smaller sectors require modeling smaller units. It monitors the battle lines and adjusts the boundaries of the rear areas like communication zones as the FEBA moves. The Theater Control Submodel assesses unit requirements, assigns replacements (weapons and personnel), and withdraws ineffective units. It calculates which airbases must be abandoned (if any) due to FEBA movement and advances units from rear areas toward combat sectors as directed by the instructions in the model or by the operators and analysts.

5. TACWAR uses the deterministic Antipotential-Potential method, a complex approach for assessing the value of each weapon toward destroying any other weapon and personnel it may engage with effect. Using a standard weapon as a reference or benchmark, analysts rate other weapons against it, building weapon-to-weapon kill matrices to determine ground weapon attrition. The engine aggregates the weapons in each sector to calculate how many opposing weapons of each type they destroy. Personnel effectiveness at operating the weapons is based on unit strength and logistics, modified by chemical warfare protective gear, training, etc. (Steve Kirin, "Executive Summary," TACWAR *Integrated Environment* (U.S. Army Training and Doctrine Command Analysis Center)).

Similar to the logistics submodel, TACWAR has a Command and Control (C2) Submodel. It overlays the sector map with a C2 grid and penalizes units' effectiveness by making fewer weapons available for ground units or generating fewer sorties if C2 is degraded. C2 effectiveness is degraded by unit casualties and by casualties to headquarters units in its chain of command. The limited Naval Submodel allows amphibious surface assault, i.e., fights to seize the beach. It treats aircraft carriers as floating airbases that generate sorties in support of the ground and air wars, and TACWAR models surface fire support like off-map artillery.

What are the strengths and weaknesses of theater-level models for force-on-force analysis? Their primary strength is that they are the only way we have that captures most of the aspects of combat we believe are so important. In this sense, they promise the greatest validity of all the approaches we have discussed so far. This validity advantage comes at the price of enormous complexity, and this is their weakness. This complexity is so great that it is very unclear (perhaps unknowable) how valid these models are at representing our theory of combat.

The reliability problem begins with the enormous amount of data and the many assumptions these models need to generate their output. Seemingly small changes in the inputs at any stage can produce disproportionate and unintended changes in the outputs. With so many inputs, we may not be able to isolate what is causing a particular result, especially if some obscure but sensitive detail of an assumption or piece of data is far upstream from the spurious result. This means that the results of an analysis can be highly sensitive to the decisions that the analysts make as they prepare the models—it can also mean that these models are subject to subtle manipulation in the hands of those who know what they are doing. For example, how will the analyst score the capability of a particular weapon? How many sorties will he permit a particular platform? How fast can tracked and wheeled vehicles transit a particular piece of terrain? What is the effect of weather on a particular suite of avionics? What is the margin of superiority required to attack? How many casualties can a unit sustain before it ceases to be effective in combat? How does an organization react when it is disconnected from its higher headquarters? How fast does combat power decline when a particular logistics node is interdicted?

The model per se does not tell us these things. Instead, it provides a platform for representing whatever values for these questions the analysts deem appropriate. It should be clear that, quite often, we have no objective way of knowing what is an appropriate value for answering such subjective questions. Thus, we can see how the services can use the same model in similar scenarios and generate different results and why they do not accept each other's analyses. We can also see why using these complex models is as much art as science.

Some have argued that an easy way around these problems exists. We could tune these theater-level models to imitate the results obtained in some real battles from World War II. The U.S. Army did this with its Concepts Evaluation Model, an Army-modified version of TACWAR that models ground war only. They ran it for the 1943 Battle of Kursk in Russia and the 1944 Battle of the Bulge in Western Europe. As one would expect, the results tracked some historical outcomes well and missed others badly. In particular, the model did a poor job of capturing the intangible factors that are so important in combat, especially morale. For example, historically, the behavior of two nearly identical Soviet units under similar combat conditions would vary widely and inexplicably; indeed, the same unit would vary its behavior from day to day. The modelers could not replicate or predict a pattern; the closest they could come was inserting random events, which was clearly unsatisfactory. In the Battle of the Bulge, German tank losses in the model were far higher than in the actual battle because it did a poor job of replicating the shock effect of an unexpected German attack on inexperienced U.S. troops.

But even if models could be made to repeat history perfectly, that approach is still not very useful. The pace of change in war is simply so great that we have little confidence that experience from previous wars is sufficiently relevant to justify setting modern theater-level models to replicate historical results. How different is combat today from combat in World War II? Also, many battle outcomes were the aggregate result of numerous low probability events and decisions. In June 1942, the Battle of Midway between the U.S. and Japanese navies was decided largely by series of tactical mistakes by commanders and intuitive decisions by small unit leaders in a sequence that is very unlikely to be repeated under any circumstance. When we tune a model to reproduce a historical result, we are in effect saying we believe those low probability events and decisions should be expected and incorporated into all our analyses and future conflicts. Plainly that is not appropriate.

Some argue that a piston-driven model based on Lanchester's Equations, such as TACWAR, is inherently incapable of representing modern warfare at the theater level. An attrition-based model cannot adequately reward maneuver let alone "effects-based warfare" that strives to paralyze an enemy's command and control of his forces and induce psychological and information warfare effects as well as physical damage. In short, an attrition-based model equates to a flawed theory of modern combat, say these critics. The counter-argument is equally simple: at some point any effect becomes discernible and affects the outcome of combat; therefore it can be included in the model—if we can agree upon the nature and the magnitude of the effect.

Like any other analytical method, a mathematical model can do no better than the theory of combat that it is intended to portray. The model cannot tell us how ground or air forces fight. It can only tell us that, given a particular theory of how they fight, a particular alternative is likely to produce a particular result. For this reason, modeling results can never be construed as a point prediction of what we can actually expect in the real world. However, we do use these models for weaker kinds of predictions, e.g., whether one alternative is likely to perform better than another in the real world.

Thus we must insist that all analyses using these force-on-force models assess the sensitivity of the results to changes in key variables and that they compare alternatives without making changes to them. The caretakers of our models should be comparing their outputs constantly with new information gleaned from actual conflict and from experiments. We should take advantage of the immense practicality of these models—once they are built, analysts can run them many, many times at little additional cost—to root out inconsistent outputs from small changes. Once again, we must never treat the output of a model, no matter how sophisticated, as something to be taken at face value. The model is never responsible for its own results; the users of the model are and they must analyze those results keeping the limitations of the model in mind.

CASE STUDY: PREDICTING THE OUTCOME OF DESERT STORM

Immediately prior to the beginning of the air campaign in the 1991 war with Iraq, Dr. Joshua Epstein of the Brookings Institution released his force-on-force analysis of the impending conflict based on his Adaptive Dynamic Model. He forecast some 16,000 American and Allied troops would be wounded and an additional 4,000 killed. His findings were widely reported on television news programs by CNN. Such losses fortunately failed to materialize. Why was Epstein so wrong? Was his mathematical model defective?

Epstein's analysis was based on the following crucial assumptions, all of which varied from the actual Gulf War:

- A short preparatory air war of one to three days;
- A direct, frontal Coalition assault on dug-in Iraqi troops;
- Healthy, supplied Iraqi troops motivated to resist.

This illustrates how assumptions and inputs can determine analytical results. Epstein assumed a short air campaign; we executed a long one. Epstein assumed a frontal attack; we executed a flanking attack. The Iraqi troops were demoralized, poorly supplied, and sick. However, the U.S. Central Command planning staff, with a vastly more complex model, obtained similar results using like assumptions. These results were one reason why the frontal attack course of action was rejected. Without actually examining his model, in light of the Central Command's results, we should suspect that the problem was in Epstein's assumptions about the campaign plan rather than his model. His model probably responded accurately to the implications of those incorrect assumptions.

Thus, we should always look carefully at the inputs when trying to understand why a model is producing a particular result. This may seem obvious, but too often the model itself receives the blame when results deviate from what is expected. Of course, there are defective models, but they are much less common than flawed assumptions and data, errors in other inputs, or mistaken theories of combat.

Exercises and Experiments

Exercises and experiments are the oldest forms of dynamic force-on-force analysis, and, at the same time, the area of most rapid development in the last few years. Exercises are performed to instill training and assess current operational concepts, tactics, procedures, and unit or crew proficiency. Experiments emphasize new concepts, tactics, and weapons, and explore possibilities for how they may be used.

Instrumented ranges are used in both exercises and experiments to increase the reliability of measurements. The various instrumented ranges for land, air, and maritime forces enable us to come closer to creating and measuring the conditions of real warfare than has ever been possible. The strength of using these ranges is that they allow us to measure criteria as close to real combat as possible. Even so, these methods fall short of the real thing in possibly critical ways. For example, most of the participants are not in fear of dying when they participate in these exercises, so we cannot capture all of the psychological dimensions of combat. Also, some important aspects of ground combat are not well-represented, such as the effects of artillery fire and air-to-ground interactions.

More importantly, exercises and experiments are elaborate and expensive, so it is difficult to repeat trials to assess alternatives. For these reasons, particular exercises and experiments tend to be one-time only events (although they may be repeated annually), and their outcome's overall reliability is low. Again we see the trade between validity and reliability in force-on-force analysis. To date, because of service preferences, the instrumented ranges have been used mainly for training rather than experiments, but this is slowly changing to explore the Revolution in Military Affairs.

Wargames

Similar to the field exercise is the wargame. Wargames generally involve less structure and more free play than exercises; there are many decision points and many branches and sequels from each choice. The annual Global War Game at the U.S. Naval War College is a good example. Players come from the Joint Staff, the services, the unified commands, defense agencies, and other government departments to examine contemporary policy and force planning issues. Wargames are like exercises in that people are directly involved, but, unlike exercises, military forces are represented abstractly and the imaginary forces operate according to game rules. Sometimes humans apply the rules; sometimes a computer performs this function.

Wargames confront individuals with a problem (for our purposes, a force-on-force problem) and require them to make decisions to solve it. Sometimes the game stops there, sometimes the players have to implement their decisions and see what would happen—at least according to the game's rules. Wargames provide practice for commanders and staffs who will actually have to make decisions like the ones modeled. Wargames help us develop a sense of how decision makers will react to a problem when they lack previous experience dealing with it.

The strength of wargames, like that of exercises and experiments, can be their high level of validity and realism when they are done well. These methods can convincingly expose individuals to situations and conditions they are unlikely to experience before they have to confront the "real thing." They are also valuable for developing new ideas and courses of action that, in turn, need further exploration. Unfortunately, wargaming's major weakness is an appearance of reality that can frequently give participants the sense that they have encountered something close to reality without really having done so; they do not faithfully replicate the real world. People come away from such experiences feeling that they learned something when they have not—at least not about the real world.

Wargames are usually too elaborate, expensive, and time-consuming to permit repeated analytical trials to test alternatives under a variety of different assumptions and conditions. Even if resources were available, wargames, like exercises and experiments, have so many decision points that it is virtually impossible to duplicate the results of a game. This is an inherent reliability problem. Nearly every juncture in a wargame involves some sort of decision, which, in turn, prescribes the path of the game while eliminating future choices. Large games involve literally millions of these decisions. Replaying the game, changing only one decision, is impossible. Thus, it is completely inappropriate to conclude from a wargame that "A" caused "B" or that the outcome of the game reliably forecasts the result of a combat situation.

In spite of these issues, wargaming is quite useful, as long as we keep the results in perspective. Wargaming can help us train participants to think through a situation well before it or something similar occurs. Wargaming provides some insights into the broad trends that might be present in a potential engagement and that deserve further analysis. For example, a wargame might reveal logistics bottlenecks, an imbalance of air power, or a real advantage if armor is used in a certain fashion. We would then seek to assess with other methods whether these findings can be substantiated or whether they are an artifact of the wargame's rules, scenario, assumptions, and participants.

We can conclude very little based only on the findings of a single wargame. Wargames, therefore, should be used primarily for training and developing hypotheses for subsequent analysis. One application of wargames, which has become increasingly common, ignores this limitation. These wargames are large, involve high-ranking personages, and receive much advance publicity. They are held to explore high-level controversial questions of intense interest to the sponsor. Some recent examples include games about Revolution in Military Affairs-type forces and systems, information warfare, and the effects of modern air power.

The problem with these wargames is that they are too high-profile to permit open acknowledgment of their weaknesses. In addition, the sponsor too often already has the conclusions that he or she seeks to prove through gaming. The pressure is too great on the participants to produce "meaningful results," often at the last minute, which then may become institutionalized. In fact, these wargames almost never produce analytically justifiable results. They do, however, serve the political purpose of giving a group of influential people a sense of ownership over a policy for which the sponsor seeks support. It is vital that we understand this distinction when we prepare for a wargame—understanding our decision maker's objective is seldom more important than during a high-level political game.

Future Issues

We noted that most existing force-on-force analysis models depend a great deal on the combat data collected from previous modern wars, particularly World War II and the Korean War. We know that technology and economics are changing the nature of military operations rapidly. Has warfare become so different today that the data from wars fought 50 years ago is almost totally irrelevant? Developments in sensor, information, materials, and communications technologies may make it possible to conduct operations in ways radically different from the past. As a result, some argue that the current generation of force-on-force models is incapable of representing the implications of the Revolution in Military Affairs and that we need a new generation of force-on-force analysis models that are not rooted in the past. In a similar vein, there are virtually no models for analyzing peacekeeping and peacemaking operations, humanitarian assistance, and similar military activities intended to shape the security environment. This was not a great problem when the primary scenario for force planning was a Soviet attack on Central Europe. Today, these operations have become the norm and we expect them to remain so. Therefore, we need new analytic methods to help us under-stand how to plan forces with these operations in mind.

The U.S. defense community is reacting to these problems in several ways. First, the Joint Staff has commissioned a new set of force-on-force models under the aegis of the Joint Analytic Model Improvement Program. The centerpiece of this effort is the development of a state-of-the-art joint, campaign-level model called the Joint Warfare System or JWARS. Scheduled for completion in 2002, JWARS is intended to address many of the difficulties assessing issues concerning the Revolution in Military Affairs. It will represent concepts such as deep maneuver; the sophisticated use of air power; the effects of advanced command, control, and communications; special operations; weapons of mass destruction; advanced logistics concepts; and missile defense.

Second, all four services and the Joint Staff have embarked on a program of field experimentation of new technologies and operational concepts. These include the Marines' SEA DRAGON; the Army's use of the National Training Center and its high technology Force XXI (a division-sized unit); the Navy's Fleet Battle Experiments; and the Air Force's Joint Expeditionary Force Experiment. In addition, the Secretary of Defense has designated the U.S. Joint Forces Command as the principal designer and integrator for an aggressive program of joint force experiments.

Third, the services have developed sophisticated battle laboratories that consist of highly detailed models and simulations to assess problems specific to each service. The battle labs for the Army, Marines, and U.S. Special Operations Command are the focal points for DoD's analysis of peacemaking and peacekeeping. The Air Force battle lab is the focal point for assessing advanced air concepts. The Navy Warfare Development Command, co-located with the U.S. Naval War College, is the Navy's clearinghouse for innovation, doctrine, new warfighting concepts, and organizing experiments to test new tactics and procedures with the numbered fleets.

JOINT WARRIOR INTEROPERABILITY DEMONSTRATIONS[6]

To test new ideas with operators and encourage the services to accelerate their use of the most promising emerging technologies, the Joint Staff annually sponsors a set of demonstrations called Joint Warrior Interoperability Demonstrations (JWIDs). Beginning in 1995, government and industry joined forces in JWIDs to demonstrate new and emerging technologies that will shape the battlefield of the future. The projects introduce off-the-shelf, new, and evolving technologies that solve command and control, communications, computer, intelligence, surveillance and reconnaissance interoperability issues for joint and combined warfighters.

A JWID is carried out over two years with an annual event each summer. Calendar Year 2000 was a Theme Year, and 2001's Exploitation Year has followed it. In the Theme Year, participants competitively assess technologies from the private sector in a military environment. The individual sponsoring combatant command for each technology demonstration and the Joint Staff establish technical criteria and specify the goals the demonstrations must achieve. In the following Exploitation Year, the "best of the best" from the Theme Year are more fully developed into an integrated evaluation. DoD, the CINCs, and the services can target these "Gold Nugget" technologies for rapid prototyping or fast-track acquisition to speed their integration into Defense Department systems.

A different military organization runs each cycle. U. S. Space Command (USSPACECOM) hosted JWID 2000 over three weeks in the summer of 2000. The JWID 2000 theme was space-based support to warfighters: integration of space forces and space-derived information with air, land, and sea forces. The demonstrations showcased global dominant battlespace awareness in combined and coalition task force settings, and the ability to unify, integrate, and expedite intelligence, surveillance, and reconnaissance support to the warfighter through a single interface. Participants also evaluated enhanced information superiority technologies in a multinational environment.

In addition to activity at their headquarters at Peterson Air Force Base, Colorado, USSPACECOM supported other JWID 2000 warfighting commands, including U.S. Pacific Command and U.S. Joint Forces Command. Numerous North Atlantic Treaty Organization nations, Australia, and New Zealand ran their own demonstrations based on scenario inputs and included command and control interoperability trials with the United States during JWIDS.

With JWIDs, DoD is creating a process that combines the operator's experience and requirements with the practical knowledge of industry and the science of the laboratory. JWIDs put low-cost, low-risk leading-edge technology in the warfighter's hands as expeditiously as possible. Also as a result of JWIDs, DoD identifies potential investment strategies toward long-range solutions to integrate programs into an enduring, interoperable system of systems.

---

6. More information on JWIDs can be found at website <http://www.jwid.js.mil>.

Summary

The necessity of doing force-on-force analysis places us in the difficult position of choosing among partially satisfactory alternatives. We cannot go to war as an analytical exercise, and we certainly cannot repeat that war to test different assumptions, systems, and forces. Instead, we have to find analytical proxies for real wars. Given the complexity of the real thing, we should not be surprised that, as we have seen in this chapter, each proxy for war so far developed is unsatisfactory in some way. Figure 8-14 summarizes the trade-offs we make between validity and reliability as we choose between models. While computers have helped us improve reliability, often dramatically, they have not resolved the fundamental uncertainties associated with our competing theories of combat.

Yet, there is no alternative to force-on-force analysis if we seek to plan military forces rationally. Decision making by procedure is entirely inadequate for planning future forces in the face of rapid changes in technology and the security environment. We are left with the art and science of mixing experience and analysis as best we can to compensate for the weaknesses of both. We value static models for their high reliability – their simplicity and their clarity. Indeed, the endless spreadsheets used in dynamic model databases are themselves static force-on-force models.

Dynamic modeling can replicate actual combat better than static models. Additionally, dynamic methods can give us powerful insights into how new systems and concepts will perform in combat. However, we must view the results of such analyses critically on the basis of experience. Whenever we encounter results that defy experience, we must inspect them in depth. To do so, we may use one type of force-on-force analysis to strengthen another. For example, results obtained from a mathematical model can be tested in a field exercise, or we can use exercise results as the basis for inputs into a mathematical model. The results of either type of analysis can be used to modify static measures. Ultimately, we must be patient and thorough, understand the strengths and weaknesses of each analytic method we encounter, and resist the impulse to surrender to the frustration from using necessarily flawed tools.

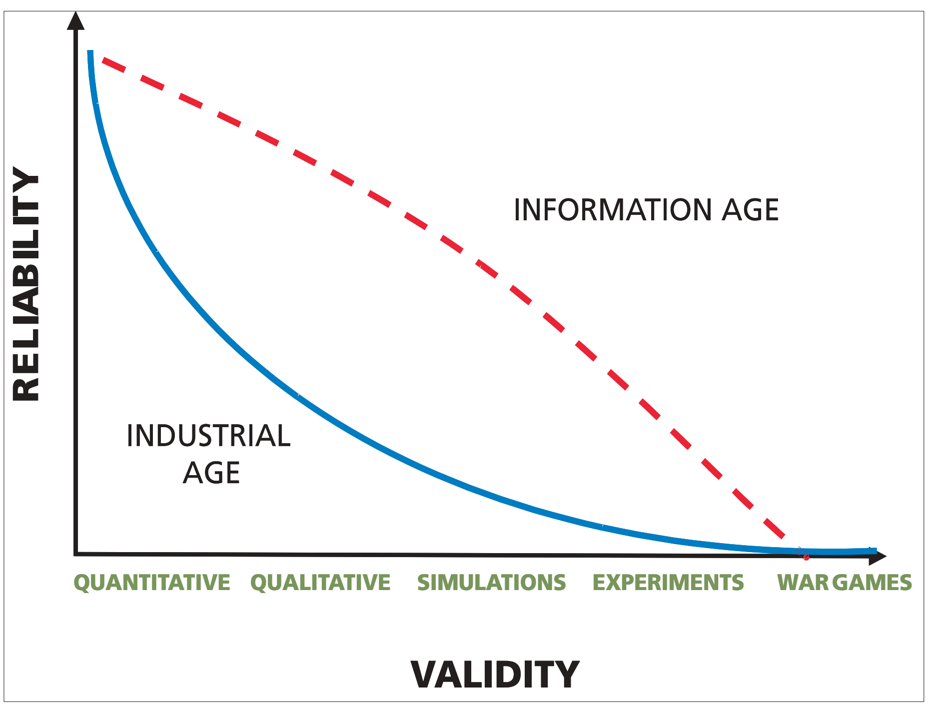


Figure 8-14. Validity and Reliability of Force-on-Force Models.